Understanding Compression Technologies for HD and Megapixel Surveillance

Understanding Compression Technologies for HD and Megapixel Surveillance

When the security industry began the transition from using VHS tapes to hard disks for video surveillance storage, the question of how to compress and store video became a top consideration for video surveillance system designers. As the industry moves from analog cameras and digital video recorders (DVR) to IP cameras and network video recorders (NVR), how to compress and store video comes into question again. When analog cameras are connected to a DVR, video compression is performed inside the recorder unit at a central location. While IP camera video compression is performed inside the camera then transmitted to the NVR in the compressed format. The centralized compression of DVRs typically meant that all cameras in the surveillance system had to use the same compression technology. IP cameras, on the other hand, have allowed for the design of hybrid systems that can use multiple compression technologies on the same system. As a result, it is critical for end-users, integrators, and system designers to have a clear understanding of the compression technologies available. Knowing when each should be used will create the best results in a system design.

There are now a wide variety of compression technologies available on the market, but no clear standard has emerged. At the same time, implementations of a particular technology may vary from one vendor to another. Often, installers think only of file and disk size and how that determines the number of days video is stored, neglecting the fact that video compression can also impact a video surveillance system design. For example, video compression technology impacts the choice of hardware for client workstations, what transmission systems can be used, and the speed, success, and efficiency of investigations.

Frame-by-Frame and Temporal Compression Technologies

There are two broad groups of compression technologies currently used in video surveillance: frame-by-frame encoding and temporal encoding. Each technology group incorporates different formats and in turn has its own tradeoffs. Understanding these differences will allow the system designer to choose the right compression technology to best meet the project's requirements and performance objectives.

Frame-by-Frame Compression

Frame-by-frame, or intra-frame, compression technologies compress video by applying a compression algorithm to each frame captured by a camera. The end result is a series of individually compressed images.

Video that is compressed using a frame-by-frame compression technology presents a number of benefits over the more complicated temporal compression technologies discussed later. First, the resulting video through frame-by-frame compression is a series of individually compressed frames that do not require information from other frames – they can be compressed and transmitted out of a camera more quickly to reduce latency. Second, because each frame acts as an independently accessible frame and is not built up from multiple frames, recorded video can be accessed more quickly. This rapid access improves investigation efficiency and can improve the forensic viability of the recorded video. In the most demanding high security situations, providing all recorded video as a series of independent video frames ensures that the video cannot be challenged due to invalid frames generated by the compression process.

The two main frame-by-frame compression technologies currently used in video surveillance are discussed in more detail in the following sections: JPEG and JPEG2000.

Figure 1 - Intra Frame Encoding = Frame by Frame

JPEG

JPEG compression is most widely used for static image compression in digital cameras and on the internet. JPEG compression is named after the Joint Photographic Experts Group and was initially introduced in 1992. Based on a compression technique known as a 'discrete cosine transform,' JPEG compression relies on blocks of pixels, typically 8x8 in size, to compress the information in an image and reduce its file size. This block-based transformation typically introduces blocking algorithms like those shown in Figure 2. These block artifacts can sometimes obscure image details when JPEG images are heavily compressed.

Figure 2 - Example image showing JPEG compression artifacts

JPEG2000

Since its introduction in 2000, JPEG2000 has gone through many revisions and updates. JPEG2000 has become a widely used standard in many different industries. For example, JPEG2000 is used in digital cinema, diagnostic medical images, document archiving, and in the capture and transmission of images from satellites and other military applications.

JPEG2000 is designed to preserve as much detail and evidence as possible within the image while greatly reducing file sizes. As a wavelet-based compression technology, JPEG2000 allows for additional compression with fewer artifacts in the image. The JPEG2000 compression process generates images that are 30 percent smaller in file size and bandwidth than a conventional JPEG image of the same visual quality, and adds additional features for effective streaming and transmission.

Two additional features of JPEG2000 compression are its ability to capture a wide dynamic range and its ability to scale to higher resolutions. Dynamic range is an important topic in surveillance because many cameras are challenged to record bright and dark areas that vary dramatically throughout the day and by season. The ability to capture dynamic range is expressed in bits. Most compression technologies capture 8-bits of dynamic range, which means it can describe 256 different intensities of light within the image. The sensors used in surveillance cameras are often capable of capturing more than 256 intensities of light and more information than even the human eye can see. JPEG2000 was designed to preserve the extra information that the sensors generate and maintain it in the compressed video. The second key feature of JPEG2000 is its ability to scale to higher resolutions, unlike technology borrowed from the consumer market. For example, JPEG2000 can scale resolutions higher than MPEG-4, which is typically limited to VGA (640 x 480 pixels) or lower resolutions. JPEG2000 is designed to scale up to extremely high resolution images and make use of its progressive compression to efficiently allow the transmission and display of those images. Information on the JPEG2000 advantage and how Avigilon has combined it with High Definition Stream Management (HDSM) for even greater results is discussed in the 'Streaming and Network Effects of Compression' section.

Temporal Compression

Temporal compression technologies rely both on compressing data within a single frame and on analyzing changes between frames. The result is a stream of video that is compressed over multiple frames rather than a series of individual frames. Typically, a temporal compression technology will attempt to store only incremental changes between frames and store whole frames only on periodic intervals. Though this technique can result in bandwidth efficiencies, it can also lead to the loss of information because the whole frame is not retained. The technologies used for temporal encoding are also often referred to as inter-frame or 'time-based' encoding because they rely on information spread out over time. The two main temporal compression technologies currently used in video surveillance are discussed in more detail in the following sections: MPEG-4 and H.264.

Figure 3 - Inter Frame = Temporal

MPEG-4

MPEG-4 compression is an umbrella term used for many different technologies defined by the Moving Picture Experts Group. Most surveillance systems implement a variant of MPEG-4 Part 2, which was introduced in 1999. However, there are many different MPEG-4 compression technologies available and few are alike. MPEG-4 compression incorporates the same basic technology as JPEG compression for reducing the file size of a digital image, but encode different types of frames in a video as a group of pictures (GOP) rather than as independent images.

A GOP is typically composed of three different frame types: I, P, and B frames. Intra-Frames (I-Frames) are complete encoded images similar to the images generated using JPEG or JPEG2000 compression. Predicted-Frames (P-Frames) are coded with reference to the previous image, which can be either another P-Frame or the previous I-Frame. Bidirectional-Frames (B-Frames) are sandwiched between I-Frames and P-Frames, and contain information on the changes calculated between the previous and subsequent frames.

Typically, MPEG-4 compression is limited to VGA resolutions and isn't commonly available for higher resolution surveillance cameras. Similar to JPEG, most implementations of MPEG-4 compression in surveillance are limited to 8-bits of dynamic range. This results in a loss of information if the camera is capable of capturing a wider dynamic range.

H.264

H.264

H.264 is the newest compression technology used in the security industry. H.264 compression is actually a variant of the MPEG-4 standard, commonly referred to as MPEG-4 Part 10 Advanced Video Coding (AVC). It uses the same basic concepts of I, P, and B Frames to encode video, but relies on more advanced coding technologies. One example is motion compensation using motion vectors to compress video to a smaller size. H.264 compression allows frames to be inserted between I-Frames in a GOP to describe the relative movement of information from a reference frame, further reducing the information required to represent video.

Another feature of H.264 that extends beyond standard MPEG-4 is the availability of de-blocking filters. De-blocking filters can smooth artifacts created by large amounts of compression. This allows systems to be configured with a higher level of compression while maintaining more detail in the images. H.264 compression is ready for use with higher resolution surveillance cameras, especially as one and two megapixel H.264 IP surveillance cameras become more widely available on the market.

Stream Size, Frame Rate, Lighting and Activity with Temporal Compression

Temporal compression technologies rely on scene changes as part of its compression methodology, and can introduce variability in the size of the compressed data stream that is generated. This variability depends on the compression being used – if it is configured to use a constant bit rate (CBR) or a variable bit rate (VBR). When configuring a system for a constant bit rate, the amount of compression applied increases as more activities occur. This can add compression artifacts to the image and degrade image quality. When variable bit rate compression is used, the size of the compressed stream is allowed to vary to maintain consistent image quality.

Variability in the size of the compressed stream presents important challenges in system design. Networks and servers should be designed for the worst case bandwidth demands. This ensures that on higher activity, a network is not overwhelmed. Storage must also be scaled for the worst case to ensure that the required retention times can be met under all conditions. Alternatively, frame-by-frame compression technologies offer a predictable (constant) compressed data stream size and therefore allows for simpler system designs.

Frame rate will also have a dramatic impact on the level of activity perceived in video by the compression technology. For example, a camera running at 30 frames per second may use a single I-Frame every two seconds and rely on changes in the scene to describe the other 58 frames in between. At this rate, the amount of change between individual frames could be very small, and substantial savings in bandwidth could be achieved by only storing scene changes for those frames. However, as the frame rate is decreased, the amount of change between frames can increase substantially. When running below 10 frames per second, there may be so much incremental change between frames that a temporal compression has little or no benefit over a frame-by-frame compression technology.
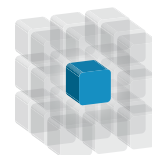
Scene lighting will also impact the ability of temporal compression algorithms to efficiently compress video. Often in low light scenes, noise within the image will be interpreted as a scene change by the compression algorithm, and cause bandwidth to increase. However, when implementing a compression technology, a camera manufacturer can optimize their motion detection algorithm to prevent the algorithm from interpreting noise in low light images as changes in the scene.

Streaming and Network Effects of Compression

By increasing camera resolution, HD and megapixel IP cameras come with their own unique challenges for storage, bandwidth, and efficient video surveillance management. These issues can be addressed by the choice of compression technology and camera resolution. Here, we will compare JPEG2000 and H.264, the most current of the frame-by-frame and temporal frame compression technologies, and review their respective strengths and weaknesses related to streaming within a network.

JPEG2000 and High Definition Stream Management

When used with high definition and multi-megapixel surveillance video, JPEG2000 can effectively and progressively compress the video and enable advance functionality for retransmitting and managing the compressed video. Avigilon has designed High Definition Stream Management (HDSM) within the Avigilon Control Center Network Video Management Software (NVMS) to deliver these key features.
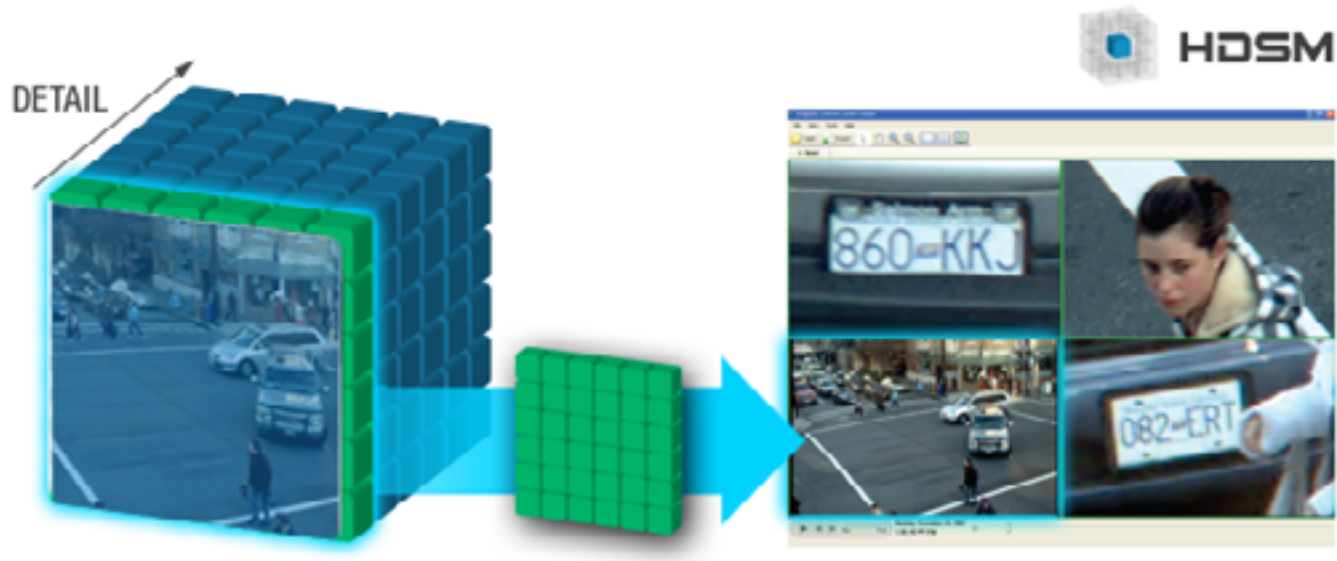
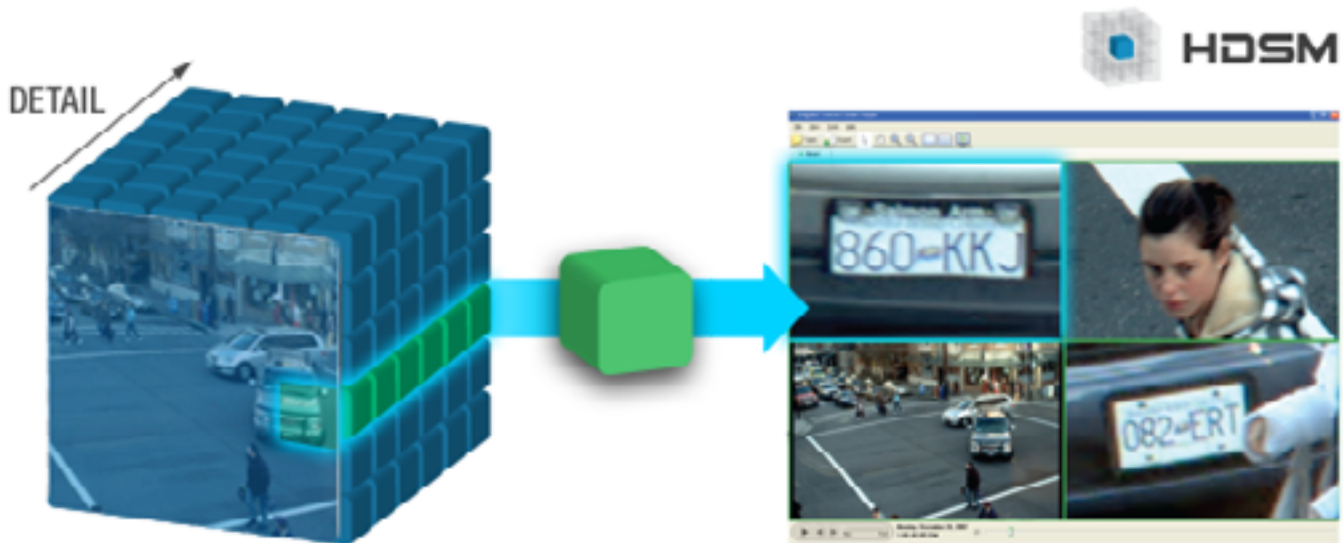Figure 4 – Streaming Situational Awareness with JPEG2000

Figure 5 – Streaming High Resolution Details with JPEG2000

HDSM gives JPEG2000 compressed video a three dimensional cube-like quality, so the video can be accessed as portions within the cube. If a low-resolution overview image is needed, only the front layer of the cube is sent, as shown in Figure 4. If a more detailed overview image is needed, additional layers of detail are sent. If a user is zooming in on a specific region to access full image detail, such as the license plate in Figure 5, HDSM will send multiple layers of that specific area for viewing. Because HDSM can dynamically access the video in layers

of detail, it can also tailor the size of the video stream being sent to the resolution of the monitor used. This results in a dramatic reduction in bandwidth between server and client compared to other compression technologies. For example, nine 5MP cameras generate 45MP of video information that is sent from camera to server. So if a connected client were viewing the video stream on a standard 1080p monitor with a total resolution of 2MP, only 2MP worth of video information is sent to the client. This allows HDSM to reduce the bandwidth between the server and client by 23 times while maintaining the ability to digitally zoom and pan within the original image at full resolution. In addition to reducing client bandwidth, HDSM also greatly reduces the processing load on the remote client. In the example, only 2MP of information is received and decompressed by the client instead of the 45MP of information.

Figure 6 – HDSM Benefits for Multiple High Resolution Cameras with JPEG2000

H.264

Video compressed with H.264 can only be streamed in the original compressed resolution and the resolution cannot progressively adapt after compression. This means that as soon as video is sent over low bandwidth connections, the ability to dynamically adapt the resolution, as is possible with JPEG2000, is no longer available. Instead of dynamically adjusting, H.264 compresses multiple streams of video within the camera at different resolutions, and sends the lower resolution stream to the client for low bandwidth live monitoring while an alternate resolution is recorded on the NVR. These additional streams add to the bandwidth transmitted from the camera but are typically much smaller than the full resolution stream being recorded to the server.

There are two important tradeoffs to remember when using H.264 multi-streaming for bandwidth management. First, when the remote client is only receiving a very-low resolution stream, it can view a scene in overview but cannot zoom in to see detail.   The second drawback is apparent when viewing recorded video. Since video management

servers are typically configured to record the higher resolution stream from the camera, there is no lower resolution stream available to view over low bandwidth connections. Without a low resolution stream, the higher resolution stream must be sent at a greatly reduced image rate.

Avigilon's HDSM technology offers a unique way of working with multiple streams of H.264 compressed video to overcome many of these trade-offs. When multi-streaming is enabled for H.264 video, HDSM will adaptively manage both a full resolution and a lower resolution stream to the NVR and viewing client. The lower resolution stream will be used for any overview streams where detail is not required, allowing efficient viewing of large numbers of H.264 compressed video streams simultaneously. When a single stream is zoomed in for more detail, that stream will be automatically sent in full resolution while other streams are kept in low resolution for an overview version of the image. This allows details to be viewed from one video stream while keeping the overall streaming bandwidth low. For example, with nine 2MP H.264 cameras and multi-streaming enabled, users can effectively generate 9 x 2MP of information for the full resolution streams and 9 x 0.3MP for lower resolution streams. When viewed on the client, the total 21MP of information is reduced to 4.4MP of information if one stream is viewed in detail and eight streams are viewed in low resolution.

Figure 7 – HDSM Benefits for Multiple High Resolution Cameras with H.264

Avigilon's HDSM also employs unique technologies to help minimize the demands placed on the client PC for decompressing H.264 video. When available, HDSM will make use of the advanced processing unit available on NVIDIA graphics cards to decompress H.264 video without using the resources of the main processor. This greatly improves the efficiency of decompression and can allow up to four 2MP (1080p) streams of video at 30 image per second to be decompressed simultaneously without using resources from the main processor. HDSM also dynamically adapts the resolution and methods used in the decompression of H.264 video to reduce the demands on the client machine. A four step control of display quality is also available to enable the manual optimization of the viewing client. You can choose to bias video display for low resolution decompression at a higher frame rate, or full resolution decompression at a lower frame rate.

Understanding Lossy and Lossless Compression and Recording

Video surveillance is fundamentally evidence, so when video is captured by a camera in its uncompressed original form, the evidence is the most pristine and contains the most information possible. When compressed, transmitted, and processed within a video surveillance system, information can be lost in different stages. When discussing compression, most people refer to either mathematically or visually lossless compression. Mathematically lossless compression refers to any compression applied to an image or video that can be completely reversed. Visually lossless compression refers to any compression that looks identical to the original image but is digitally compressed in a way that can not be seen by the human eye. Though different compression technologies can support lossless compression, it requires an end-to-end system to manage and properly store losslessly compressed video effectively. The Avigilon Control Center NVMS along with JPEG2000 and HDSM is one solution that effectively manages visually lossless video.

Information can be lost in a video surveillance system because of how it is recorded. Anyone who has worked with the analog DVRs first introduced on the market has experienced lossy recording. With these devices, the images streamed live to analog monitors were the full 540 TVL analog images captured by the analog camera, but what the DVR typically recorded was a highly compressed low resolution 352 x 288 CIF (Common Intermediate Format) image. Similarly, modern IP cameras can create multiple streams at different resolutions and some systems will display the higher resolution live stream to the client while recording a more compressed or lower resolution stream to the server. To ensure that the live video is the same quality as the recorded video, it is important to use a system that supports lossless recording. With lossless recording, systems like the Avigilon Control Center NVMS are designed to ensure that what is viewed live is the same high quality video that is recorded to the server.

Choosing the Right Compression Technology

Choosing the right camera for each point in a system is critical to a successful video surveillance system design. The first factor to consider will always be resolution. Available online tools that include pixels on target calculators are instrumental in helping select the most appropriate camera. Once resolution has been determined, it is important to select a compression technology suited to the application. Resolution, frame rate, activity level, and investigative needs will all influence the selection of compression technology. If the end-user's goal is to cover a larger scene by using a high resolution camera above 2MP, then a frame-by-frame compression technology like JPEG2000 may be the only effective option. As resolution increases, the benefits of JPEG2000 and HDSM increase. For scenes on a site that require large multi-megapixel IP cameras, JPEG2000 compression is required to manage the high resolution information effectively.

As resolution increases and frame rate decreases, the benefits of temporal compression technologies like H.264 diminishes. But, for smaller scenes that require a higher frame rate and use a camera with 2MP resolution or less, a temporal compression may be the most effective option. For low resolution high frame rate scenes, a temporal compression technology like H.264 can help minimize the required network bandwidth and storage. For scenes of moderate activity with high frame rates, H.264 compression can offer a significant advantage.

Compression Technology Selection by Application

Application	Resolution	Image Rate	Compression Technology
Parking Lot	16MP	3	JPEG2000
Cafeteria	5MP	7	JPEG2000
Lobby	3MP	7	JPEG2000
Doorway	2MP	15	H.264
Hallway	1MP	15	H.264
Casino	1MP	30	H.264